Clinical and Translational Research

Delineation of fatty acid metabolism in gastric cancer: therapeutic implications

Yu Fu, Bin Wang, Peng Fu, Lei Zhang, Yi Bao, Zhenzhen Gao

Abstract

BACKGROUND
The prognosis of gastric cancer is extremely poor. Metabolic reprogramming involving lipids has been associated with cancer occurrence and progression.

AIM
To illustrate fatty acid metabolic mechanisms in gastric cancer, detect core genes, develop a prognostic model, and provide treatment options.

METHODS
Raw data from The Cancer Genome Atlas and Gene Expression Omnibus databases were collected and analyzed. Differentially expressed fatty acid metabolism genes were identified and incorporated into a risk model based on least absolute shrinkage and selection operator regression analysis. Then, patients from The Cancer Genome Atlas were assigned to high- and low-risk cohorts according to the mean value of the risk score as the threshold, which was verified in the Gene Expression Omnibus database. Relationships between chemotherapeutic sensitivity and tumor microenvironment features were assessed.

RESULTS
An integrated evaluation was performed in this study. Fatty acid metabolism-related genes were used to construct the risk model. Patients classified into the high-risk cohort were considered to be resistant to chemotherapy based on results of the “pRRophetic” R package. Patients in the high-risk cohort were associated with type I/II interferon activation, increased inflammation level, immune cell infiltration, and tumor immune dysfunction based on the exclusion algorithm, indicating the potential benefit of immunotherapy in these patients.

CONCLUSION

We constructed a fatty acid-related risk score model to assess the comprehensive fatty acid features in gastric cancer and validated its vital role in prognosis, chemotherapy sensitivity, and immunotherapy.

INTRODUCTION

Gastric cancer (GC) is associated with poor overall survival and is one of the most common types of fatal cancer worldwide[1, 2]. Regional differences in the incidence of GC have been reported. Globally, the 5-year overall survival rate is reported to be less than 25%, given the late-stage presentation, metastasis, tumor heterogeneity, and the intrinsic therapeutic resistance of GC[3]. Joshi et al investigated GC biology and found that, compared with cases in Western countries, GC cases in eastern Asia were characterized by lower proportions of signet ring cell histology[3]. However, the 5-year overall survival rates remain low. Identifying the novel mechanisms underlying GC is of critical importance to develop treatment strategies that can be used in clinical practice.

In general, cancer is characterized by the interactions between solitary metabolic features and the surrounding complex tumor environment[4]. Carbohydrates, lipids, and amino acids are the main nutrients in cancer cells, among which lipids are crucial mediators in energy metabolism and signal transduction. Importantly, fatty acids are the basic components of diverse types of lipids[5, 6]. Furthermore, metabolic
reprogramming has been identified as a critical hallmark of tumor cell proliferation and differentiation. Ma et al described various transcription factors controlling gene expression programs including fatty acid metabolism (FAM)-related genes\(^7\). Khan confirmed that SIRT6 could regulate fatty acid transport by suppressing PPAR signaling\(^7\), thereby promoting tumor progression. Luo et al demonstrated that the critical role of fatty acid oxidation in CD8+ T cell memory formation, which is crucial in the tumor microenvironment\(^8\). García et al found that elevated levels of fatty acid may accelerate cancer progression through endoplasmic reticulum stress by modulating the forkhead box (FOX)O3-FOXM1 axis\(^9\). In addition, Zhang et al demonstrated that fatty acid-induced CD36 expression promoted the progression of GC by activating the nuclear factor-κB pathway and direct binding of CD36 at S468 and T470\(^10\). Adenosine triphosphate citrate lyase, acetyl-CoA synthases, acetyl-CoA carboxylase, and fatty acid synthase are crucial enzymes in fatty acid synthesis\(^11\). Although various molecules have been identified in GC metastasis, uncovering novel links between fatty acids and the tumor environment are necessary for a comprehensive understanding of GC and will contribute to accurate diagnosis, prognosis prediction, and recurrence risk and metastasis assessments in patients with GC.

In this study, we constructed a novel prognostic risk model for GC based on FAM-related genes. We examined the mechanisms underlying FAM in GC, the relationship between the risk of GC and tumor microenvironment characteristics, and treatment strategies for GC.

**MATERIALS AND METHODS**

Data download and analysis

Raw data for GC were downloaded from The Cancer Genome Atlas (TCGA) project (https://portal.gdc.cancer.gov/) and the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). Transcription profiling data from TCGA-stomach adenocarcinoma project were collected in the stomach carcinoma-counts workflow type. Simple nucleotide variation data were also retrieved as masked somatic
mutation data from TCGA-GDC portal. Clinical information of patients with GC including age, sex, tumor stage, tumor grade, T/N/M (tumor, node, metastasis) stages, and survival information was also obtained. The microarray data from the GSE84437 dataset in the GEO GPL6947 platform were also acquired. The annotation platform was implemented to convert the Entrez Gene identifiers (IDs) of each case to gene symbols.

Detection of differentially expressed genes and functional enrichment analysis between normal and cancer samples

Generally, FAM-related genes were acquired, as previously described. Differentially expressed genes (DEGs) in normal and tumor cases were identified using the “limma” statistical package in R (version 4.1.3; R Foundation for Statistical Computing, Vienna, Austria). A false discovery rate (FDR) <0.05 was considered the cutoff value. Conversion from Entrez Gene IDs to gene symbols was processed using the “org.Hs.eg.db” R package. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were conducted to detect the biological characteristics using the “clusterProfiler” R package. Finally, enrichment analysis results were visualized using the “enrichplot” and “ggplot2” packages. A P value less than 0.05 (P < 0.05) was considered statistically significant.

Establishment and validation of prognostic risk model

Clinical samples from TCGA database were defined as the training cohort, and the GSE84437 dataset from the GEO database was defined as the validation cohort. We combined the expression level of FAM-related genes with clinical information using the patient IDs. Genes related to survival were based on DEGs and analyzed using univariable cox regression, in line with the validation cohort. Then, genes with statistically significant values (<0.05) were selected for further analysis. Prognostic genes were generated through least absolute shrinkage and selection operator (LASSO) cox regression analysis using the “glmnet” package, and a prognostic model was constructed. The risk score was calculated as described previously. All samples were segmented into high-risk and low-risk cohorts based on the median value of the risk scores. Kaplan-Meier analysis and the log-rank test were applied to determine the
survival difference between low- and high-risk score groups. Finally, the "survivalROC" package was used to assess the predictive accuracy of the prognostic risk model. All data were verified in the validation group.

Principal component analysis and gene set variation analysis. The Ggplot2 and limma R packages were utilized to conduct principal component analysis (PCA) of gene expression profiles to construct a risk model from TCGA database. Accordingly, PCA was performed on FAM-related gene expression. Two-dimensional diagrams were then visualized to present the results. In addition, gene set variation analysis (GSVA), a non-parametric method implementing gene sets from the molecular signature (https://www.gsea-msigdb.org/gsea/msigdb) database was applied to identify the biological function and pathways between high- and low-risk groups. An FDR < 0.05 was considered statistically significant.

Construction of protein-protein interaction network. The expression profiles of DEGs between high- and low-risk models were analyzed. The gene symbols were merged into an online STRING database (https://cn.string-db.org/) to establish the protein-protein interaction (PPI) network with interaction scores > 0.7. The PPI network was then presented using Cytoscape software (version: 3.9.2, The Cytoscape Consortium, San Diego, CA, USA). Next, CytoHubba (version 0.1) was applied to generate hub genes from all the DEGs. Based on the previous genes, GO and KEGG enrichment analyses were used to illustrate the biological functions of the identified hub genes. Finally, the differences in the degree of immune cell infiltration based on the expression of hub genes were determined.

Statistical analysis. R packages including "pRRophetic" and "rms" were applied to demonstrate the predictive value of the clinical features. The Wilcoxon rank sum and Kruskall-Wallis tests were used to perform group comparisons. Kaplan-Meier analysis was used to assess survival. Univariable and multivariable regression models were applied to identify independent factors related to overall survival in GC. Receiver operating
characteristic (ROC) curves were drawn to verify the value of the risk model. All analyses were performed using R (version 4.1.3). Statistical significance was defined as \( P < 0.05 \).

**RESULTS**

Functional analysis of FAM-related genes between normal and cancer samples from TCGA database

We conducted differential expression analysis of FAM-related genes using the limma R package. In total, 113 DEGs with \( |\log FC| = 1.5 \) and \( FDR \leq 0.05 \) were identified in TCGA cohort, of which, 60 genes were upregulated, and 53 genes were downregulated, as shown in Figure S1A. Thereafter, functional analysis was conducted for the DEGs. FAM-related genes were enriched in fatty acid metabolic, acyl-CoA metabolic, and fatty acid biosynthetic processes in GO as well as KEGG analyses (Figure S1B).

Construction and validation of the prognostic risk model in GC Clinical cases from TCGA database cohort were included in the training set. Univariable cox regression analysis was conducted among the 113 DEGs related to FAM. Overall, statistically significant links to GC prognosis were detected in 17 genes \( (P < 0.05; \text{Figure 1A}) \). Next, the relationship among the 17 FAM-related genes was calculated using a correlation test (Figure 1B). In total, FAM-related gene mutations occurred in 65 patients, equating to a frequency of 16.5\% (Figure 1C). The cyclooxygenase 1 (PTGS1) gene had the highest mutation frequency among patients with GC. Furthermore, a co-occurrence mutation was demonstrated between PTGS1 and other FAM-related genes. LASSO regression analysis was applied to generate the risk score model. Collectively, six genes (monoamine oxidase A (MAOA), CD36, fatty acid elongase 2 (ELOVL2), alcohol dehydrogenase 4 (ADH4), gamma-glutamyltransferase 5 (GGT5), ELOVL4) were included in the risk model (Figure 1D-E).

The risk score for each patient was calculated based on the model and the patients were categorized into high- and low-risk groups (Figure 1F-G).
Correlation between risk model score and clinical features

We defined TCGA-GC cases as the training cohort and GEO cases as the validation cohort. In total, 339 patients with GC were recruited from TCGA database. Of these, 169 were considered to have high risk and 170 were considered to have low risk using the median risk score value. The relationship between clinical characteristics and risk model score was investigated, including age, sex, tumor grade, and T/M/N stages. No significant differences between groups were detected according to sex, M stage, and tumor grade (Figure 2A-C). Nonetheless, higher risk scores were associated with higher T (tumor invasion) and N (lymphoid metastasis) stages and with older age (Figure 2D-F). Furthermore, patients with higher risk scores had lower survival rates (Figure 2G), which was validated in the validation group based on patient data derived from the GEO database (GSE84437) (Figure 2H), including 231 patients classified as high-risk cohort and 202 patients classified as low-risk cohort. Moreover, time- and variable-dependent ROC analysis was performed to verify the results (Figure 2I-J). Univariable and multivariable regression models were constructed according to age, sex, tumor grade, tumor stage, and risk score. The risk score was an independent prognostic predictor of overall survival (Figure 2K-L).

Establishment of predictive nomogram in patients with GC

We constructed a nomogram involving tumor grade, T/N/M stages, sex, tumor stage, and risk score to predict overall survival in patients with GC (Figure 3A). The calibration curves at 1-, 3-, and 5 years demonstrated that the nomogram was relevant to overall survival in patients with GC (Figure 3B). Multivariable regression analysis revealed nomogram score as an independent index (Figure 3C), which may be a superior indicator than other clinical features used in clinical practice (Figure 3D).

A FAM-related model predicting response to chemotherapy and GSVA in high- and low-risk groups

The FAM-related risk score is associated with overall survival in patients with GC; therefore, demonstrating the association between risk score and chemotherapy response is important. The “pRRophetic” R package was used to calculate the half maximal
inhibitory concentration, which was applied to evaluate the sensitivity of the risk score and response to 5-Fluorouracil (5-FU). As shown in Figure 4A-B, the high-risk group was more sensitive to 5-FU in TCGA cohort. Furthermore, a negative relationship was noted between risk score and progression-free survival, indicating that the high-risk group may be more sensitive to chemotherapy (Figure 4C). However, its exclusive molecular type and tumor mutation burden facilitated the progression of tumor by activation of focal adhesion pathways, transformation of growth factor-beta signaling, and contraction of vascular smooth muscle (Figure 4D), ultimately shortening survival. These findings are consistent with previously reported results[12]. In line with previous studies, GSVA was performed to investigate the biological features between high- and low-risk groups by implementing “c2.cp.kegg.v7.2” in R Studio acquired from the Molecular Signature Database. As shown in Figure 4D, DNA replication, mismatch repair, and steroid biosynthesis were highly enriched in the low-risk group. Some molecules incorporating the transcription activator BRG1 (SMARCA4) and ataxia telangiectasia mutated (ATM) proteins are reportedly correlated with immune activation (Figure 4E-F). Our results indicate that ATM mutation is associated with a positive response to immunotherapy.

Immune differences between high- and low-risk groups

Regarding immune cell infiltration, distinctive enrichment of macrophage M2 was detected in the high-risk group, in line with the previous survival analysis (Figure 5A). Additionally, para-inflammation, T cell co-inhibition, T cell co-stimulation, type I interferon (IFN) response, and type II IFN response were activated in the high-risk group (Figure 5B). Different immune subtypes were identified (Figure 5C) and validated using the TIDE (Tumor Immune Dysfunction and Exclusion) algorithm.

Functional enrichment analysis of FAM-related DEGs in the low- and high-risk score groups based on PPI

The expression profiles of DEGs were collected and analyzed to establish a PPI network using a STRING online database, which was then visualized using Cytoscape. As shown in Figure 6A, upregulated genes in the high-risk group are labeled in red, and
those in the low-risk group are labeled in blue. Thereafter, the CytoHubba plug-in package was applied to generate the hub genes of FAM-related DEGs. In general, 10 genes (integrin beta (ITGB)-3, integrin alpha (ITGA)-1, fibronectin (FN)-1, fibroblast growth factor (FGF)-2, von Willebrand factor (VWF), ITGA5, ITGA4, ITGA8, thrombospondin (THBS)-1, ITGA9) were selected (Figure 6B). Functional enrichment analysis was performed using the “GOplot” R package, and these genes were involved in the integrin-mediated signaling pathway, cell-substrate adhesion, cell-matrix adhesion (Figure 6C), extracellular matrix-receptor, phosphoinositide 3-kinase-Akt signaling pathway, focal adhesion, and regulation of actin cytoskeleton (Figure 6D). Finally, survival and correlation analyses were conducted to determine the key regulators including FGF2 and THBS1, which were verified to be negatively associated with survival, consistent with previous studies (Figure 6E-F).

**DISCUSSION**

Cancer proliferation and progression are associated with metabolic reprogramming\textsuperscript{[13-16]}. The metabolism alteration theory was first proposed by Otto Warburg in the 1920s and is known as the “Warburg phenomenon,” characterized by increased glycolysis in the context of an environment lacking oxygen\textsuperscript{[17-19]}. Lactic acids generated from glycolysis favored tumor evasion and metastasis and restrained the antitumor immune effect\textsuperscript{[6,7,20]}. In addition, glucose was converted into nicotinamide adenine dinucleotide phosphate by cancer cells to cope with the complex tumor environment thereby promoting fatty acid synthesis\textsuperscript{[21-23]}. Meanwhile, the medium pyruvate generated from glycolysis was manipulated by the cancer cells to enter a truncated tricarboxylic acid cycle, thereby creating acetyl-CoA that was utilized for long-chain fatty acid synthesis. Collectively, this process highlights the involvement of FAM in cancer development and metastasis.

Few studies have identified the role of FAM in GC. Cui et al concluded that abnormal FAM had discernible effects on GC growth\textsuperscript{[24,25]}. Aberrant expression of FAM-related genes was associated with chemotherapeutic drug resistance and recurrence. Sterol O-
acyltransferase (SOAT)1 was highly expressed in GC tissues and negatively related to GC prognosis via regulation of sterol regulatory element-binding protein (SREBP)-1 and SREBP2 expression, which propelled lymph-angiogenesis through increased expression of vascular endothelial growth factor C[26]. However, other innovative mechanisms of FAM in GC warrant further investigation. In this study, we explored and elaborated on the relationship between FAM-related genes and GC using TCGA and GEO databases. FAM-related genes incorporated into the present study were obtained as previously described. We conducted a differential analysis to identify DEGs. A total of 113 FAM-related genes were screened to detect genes involved in GC progression, tumor microenvironment involvement, and prognosis. In total, 17 genes were included to establish a predictive/prognostic risk model using cox regression and then verified using LASSO regression analysis. TCGA cohort was considered as the training set and the GEO cohort as the validation set. The overall survival prediction of the risk score (each sample from TCGA cohort had a risk score, which was used to classify patients into high-risk and low-risk cohorts) was constructed in TCGA database and verified in the GEO database. The risk score generated from this model was confirmed as an independent and accurate prognostic index. Chemotherapeutic sensitivity between high-risk and low-risk groups was analyzed to evaluate the predictive value of the risk model in drugs screen in GC treatment. The increased risk score was positively associated with 5-FU resistance, consistent with previous studies[27]. In parallel, patients in the high-risk score group had reduced survival rates, suggesting that the risk model based on FAM-related genes may be applied to determine the optimal treatment (chemotherapy vs radiation therapy) for patients with GC. Moreover, patients from the high-risk score group were enriched in stroma immune activation, indicating cytotoxic drug resistance, consistent with previously reported studies[28-30]. Furthermore, high-risk score patients exhibited activation of the type I IFN response, T cell co-inhibition, type II IFN response, and para-inflammation function, demonstrating that patients with high-risk scores may benefit from immunotherapy, on account of the TIDE analysis. Altogether, the
predictive/prognostic risk model established in this study may be valuable in clinical practice.

Significant differences in gene expression were detected between the high-risk and low-risk core cohorts. Therefore, we further evaluated FAM-related genes from the risk model. Expression of fatty acid binding protein 4 (FABP4), THBS1, and other marker levels were significantly different between the two cohorts. Various studies have suggested that pro- or antitumor manifestation of FABP4 was associated with various cancer types\textsuperscript{[30-34]}. Further, FABP4 reportedly promotes breast cancer cell invasion by regulating the interleukin-6/signal transducer and activator of the transcription-3 axes. Moreover, its inhibition by BMS309403 (inhibitory molecule of FABP4) can prevent the proliferation and metastasis of ovarian cancer. Additionally, increased expression of FABP4 has been associated with a worse prognosis in GC, although Chen \textit{et al} came to a contrary conclusion\textsuperscript{[35]}. On the contrary, increased FABP4 expression prohibited tumor progression and metastasis in hepatocellular carcinoma. Accordingly, THBS1 was identified as a hub gene that can serve as an indicator for GC diagnosis. However, its predictive and prognostic role in patients with GC remains elusive. Consequently, demonstrating the role of FABP4 and THBS1 in GC in future studies is important and requires urgent attention.

**CONCLUSION**

We constructed a fatty acid risk score model to assess the intact fatty acid features in GC. This risk model combined with clinicopathological characteristics, prognosis, chemotherapy sensitivity, and immune cell function, can be used to develop personalized treatment strategies for patients with GC. This study may sharpen and enrich the rationale used in prediction models to optimize the care of patients with GC.

**ARTICLE HIGHLIGHTS**

*Research background*
the interactions between solitary metabolic features and the surrounding complex tumor environment characterize cancer. Fatty acids are the essential components of diverse types of lipids, which are crucial mediators in energy metabolism and signal transduction. Although various molecules have been identified in gastric cancer (GC) metastasis, uncovering novel links between fatty acids and the tumor environment is necessary to comprehensively understand gastric cancer.

**Research motivation**

We constructed a novel prognostic risk model for GC based on fatty acid metastasis (FAM)-related genes. We examined the mechanisms underlying FAM in GC, the relationship between the risk of GC and tumor microenvironment characteristics, and treatment strategies for GC.

**Research objectives**

We want to construct novel links between fatty acids and the tumor environment to contribute to accurate diagnosis, prognosis prediction, and recurrence risk and metastasis assessments in patients with GC.

**Research methods**

Data download and analysis; Detection of differentially expressed genes and functional enrichment analysis between normal and cancer samples; Establishment and validation of prognostic risk model; Principal component analysis and gene set variation analysis; Construction of protein-protein interaction network

**Research results**

Functional analysis of FAM-related genes between normal and cancer samples from TCGA database; Construction and validation of the prognostic risk model in GC; Correlation between risk model score and clinical features; Establishment of predictive nomogram in patients with GC; A FAM-related model predicting response to
chemotherapy and GSVA in high- and low-risk groups; Immune differences between high- and low-risk groups; Functional enrichment analysis of FAM-related DEGs in the low- and high-risk score groups based on PPI.

**Research conclusions**
A fatty acid risk score model to assess the intact fatty acid features in GC was constructed. This risk model combined with clinicopathological characteristics, prognosis, chemotherapy sensitivity, and immune cell function.

**Research perspectives**
We established a prognostic risk model using data collected from The Cancer Genome Atlas database, explored the function of the risk model, and identified the relationship between the risk model and clinical features.
<table>
<thead>
<tr>
<th>PRIMARY SOURCES</th>
<th>ORIGINALITY REPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chengsheng Ding, Zezhi Shan, Mengcheng Li, Hongqi Chen, Xinxiang Li, Zhiming Jin. &quot;Characterization of the fatty acid metabolism in colorectal cancer to guide clinical therapy&quot;, Molecular Therapy - Oncolytics, 2021</td>
<td>55 words — 1%</td>
</tr>
<tr>
<td>2. <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a></td>
<td>38 words — 1%</td>
</tr>
<tr>
<td>3. <a href="http://www.frontiersin.org">www.frontiersin.org</a></td>
<td>27 words — 1%</td>
</tr>
<tr>
<td>4. <a href="http://www.hindawi.com">www.hindawi.com</a></td>
<td>24 words — 1%</td>
</tr>
<tr>
<td>5. Lilan Yi, Guowu Wu, Longhua Guo, Xiaofang Zou, Ping Huang. &quot;Comprehensive Analysis of the PD-L1 and Immune Infiltrates of m6A RNA Methylation Regulators in Head and Neck Squamous Cell Carcinoma&quot;, Molecular Therapy - Nucleic Acids, 2020</td>
<td>18 words — &lt; 1%</td>
</tr>
<tr>
<td>6. downloads.hindawi.com</td>
<td>13 words — &lt; 1%</td>
</tr>
<tr>
<td>7. pubmed.ncbi.nlm.nih.gov</td>
<td>13 words — &lt; 1%</td>
</tr>
</tbody>
</table>